Fine tuning cognitive assessment in the elderly: The benefits of an online test battery

Avital Sternin\*, Jessica A. Grahn, Adrian M. Owen

\* avital.sternin@uwo.ca

AS, JAG & AMO  
Brain and Mind Institute  
Department of Psychology  
University of Western Ontario  
London, Ontario  
Canada  
N6A 3K7  
519 661 2111 x86057

**Funding**

This project was supported by the National Science and Engineering Research Council [418550/2012]; and the Canada Excellence Research Chairs Program [#215063].

**Acknowledgements**

Special thanks to residents of Christie Gardens Apartments and Care in Toronto, Ontario and to Grand Wood Park in London, Ontario for their willingness to participate in our research. Thank you to Emily S. Nichols and Alexandra A. Pearce for their help with data collection.

**Abstract**

**Background and Objectives:** Assessing an individual’s cognitive capacity is an important part of caring for the elderly. We sought to determine whether a novel online cognitive test battery could differentiate individuals with ambiguous cognitive scores; specifically, we asked which combination of online tests, best categorizes individuals with ambiguous MoCA and MMSE scores.

**Research Design and Methods:** 52 elderly participants completed 12 online tests on a tablet computer, a MoCA, and a MMSE.

**Results:** The MoCA categorized 73% of participants as impaired or unimpaired. The addition of a single online test increased categorization to 94%. A multiple regression identified two other tests that best predicted MoCA scores. The combination of scores from the three identified tests were highly correlated with MoCA scores. A regression also identified two tests that best predicted MMSE scores, but the categorization analysis was not performed because of a ceiling effect in MMSE scores.

**Discussion and Implications:** The addition of a single online test to the MoCA improved categorization of individuals with ambiguous scores and a short battery of three CBS tests is a viable alternative to the paper-pencil tests currently used to monitor cognitive changes in older adults. This online testing battery may have significant consequences for care and quality of life in the aging population.

Keywords: MMSE, MoCA, cognitive impairment, cognitive screening measures, aging

Translational Significance: Our results demonstrate that a novel online cognitive test battery improves the accuracy of currently used methods for monitoring cognitive functioning in older adults, while reducing administrator burden and practice effects. This online testing battery may have significant consequences for care and quality of life in the aging population.

**Background and Objectives**

Assessing cognitive capacity is important to caring for the elderly, as such assessments determine the level of care an individual requires. Full assessments are difficult and time-consuming to administer, therefore shorter versions are used to understand an individual’s capacity and efficiently follow cognitive changes over time. Long-term monitoring of cognitive abilities is important for identifying fluctuations that may require modifications to an individual’s care plan. One commonly used test is the Mini-Mental State Examination (MMSE), developed to evaluate psychiatric patients (Folstein, Folstein, & McHugh, 1975) and widely used in aging populations. The Montreal Cognitive Assessment (MoCA)(Nasreddine et al., 2005) is also popular, largely due to its brevity (under 10 minutes) and its greater sensitivity to mild cognitive impairments than the MMSE (Gluhm et al., 2013).

One shortfall of the MoCA and the MMSE is ambiguity about determining threshold (or ‘cut off’) scores (Nasreddine, Phillips, & Chertkow, 2012) which affects whether individuals are classified as cognitively impaired or unimpaired. For example, the recommended MoCA threshold may be too high for aging populations (Damian et al., 2011; Gluhm et al., 2013; Malek-Ahmadi et al., 2015). A recent study found that an online cognitive battery of two tests improved the classification of individuals with ambiguous MoCA scores (Brenkel, Shulman, Hazan, Herrmann, & Owen, 2017), suggesting that these newer tests may more accurately classify cognitive abilities than traditional approaches.

We used the Cambridge Brain Sciences (CBS) test battery (cambridgebrainsciences.com) (Hampshire, Highfield, Parkin, & Owen, 2012) to extend this preliminary investigation, examining whether a more extensive battery of 12 cognitive tests would improve identification of individuals with cognitive impairments relative to the MoCA and the MMSE. The CBS test battery’s novel approach to cognitive testing is based on standard neuropsychological tests. However, the tests are computerized and available online, with comprehensive instructions, practice trials, and ‘guided learning’ videos to ensure that individuals can complete them without an examiner being present. Additionally, difficulty levels scale with ability, and test items are randomized, creating a unique set of stimuli for the participant every time the test is taken. Here we asked which CBS test, or combination of tests, best categorizes individuals with ambiguous MoCA and MMSE scores.

**Research Design and Methods**

**Subjects**

Participants over the age of 50 were recruited from retirement homes in Toronto and London, Ontario. Any participant who was unable to provide informed consent, or understand task instructions, was excluded. Fifty-two participants (43 female) participated. Possibly because of the location of the retirement homes, the sample was highly educated. All but one earned high school diplomas, 24 earned postsecondary degrees, and 16 earned postgraduate degrees. The study was approved by the University of Western Ontario Research Ethics Board.

**Procedure**

All participants were asked to complete the 12 online tests from the Cambridge Brain Sciences (CBS) battery in random order (descriptions are in the Supplementary Materials) Each task was presented on a touchscreen tablet computer and was preceded by instructions and practice trials. Researchers offered clarification if necessary. Participants took breaks between tasks to prevent fatigue. Afterward, the MoCA (version 7.1 English) (Nasreddine et al., 2005) and MMSE (Folstein et al., 1975) were administered in interview format, always by the same person (AS). Participants also completed a paper demographic questionnaire.

**Results**

Fifty-two older adults (average age = 81 years, 62-97 years) were asked to complete 12 CBS tests, the MoCA, and the MMSE. Two participants did not complete all 12 tasks due to fatigue and loss of interest, thus 50 participants’ scores were analysed. MoCA scores ranged from 12-30 (mean=24.6) and MMSE scores ranged from 16-30 (mean=27.7). A summary of all task scores is in Table 1.

\*\*\* Table 1 about here please\*\*\*

A step-wise multiple regression showed that MoCA scores were best predicted by two CBS tests: Feature Match and Odd One Out (R2=0.65). Age did not significantly predict any variance over and above these tests. Alone, age predicted 22% of the variance in MoCA scores (R2=0.22). Another step-wise multiple regression showed that MMSE scores were best predicted by Feature Match and Grammatical Reasoning (R2=0.38). Age did not explain a significant amount of variance over and above the task scores. Alone, age predicted 8% (R2=0.08) of the variance in MMSE scores.

A third regression showed that level of education did not explain a significant amount of variance in MMSE or MoCA scores, although this may be due to overall high educational levels and the ceiling effect seen in MMSE scores (see Figure 1).

Participant scores were split into three categories based on MoCA scores (See Figure 1): unimpaired (n=25) MoCA score ≥26, borderline cognitive impairment (n=14) MoCA score 23-25, and impaired (n=12) MoCA score ≤ 22, based on thresholds from previous literature (e.g. Damian et al., 2011; Gluhm et al., 2013; Malek-Ahmadi et al., 2015). The ceiling effect precluded performing this analysis on MMSE results.

\*\*\* Figure 1 about here please \*\*\*

To replicate Brenkel et al. (2017), each participant in the borderline group was reallocated to either the impaired or unimpaired groups based on their CBS test scores as follows: If the participant’s score on one of the 12 tasks was less than or equal to the average score of the impaired group (on that task) they were categorized as impaired. If their score on a particular task was greater than or equal to the average score of the unimpaired group (on that task) they were classified as unimpaired. If their score fell between the average scores of the impaired and unimpaired groups they remained classified as borderline. This procedure was conducted for each individual CBS test as well as all possible combinations. When multiple tests were used, participants were only categorized if categorization was consistent across all tests in the combination.

With the MoCA alone, 73% of participants were classified as impaired or unimpaired. Adding one CBS test (Spatial Planning) increased categorization the most (94%), leaving only 3 participants in the borderline group. This was not because Spatial Planning was the most difficult test, as the equally difficult Spatial Span test left 5 participants in the borderline group. Test difficulty was determined from an unrelated study with scores from 327 participants age 71-80 (see Supplementary Figure 1).

Participants’ scores on the three tests identified in the stepwise regressions (Feature Match, Odd One Out, and Spatial Planning) were converted to z-scores and averaged to create a composite. This composite score strongly correlated with MoCA scores *r*=0.74 (*p*<0.001), and was slightly less correlated with MMSE scores (*r*=0.55, *p*<0.001; see Figure 1).

**Discussion and Implications**

In this study, we found that tests from the CBS online cognitive battery successfully identified cognitive impairment when the MoCA or MMSE returned ambiguous scores. Specifically, Feature Match and Odd One Out tasks best predicted MoCA scores. A further categorization analysis showed that considering one computerized test, Spatial Planning, in conjunction with the MoCA, classified 94% of participants as impaired or unimpaired (compared to 73% with the MoCA alone). Better classification of individuals with ambiguous scores has implications for treatment and quality of life. We were unable to perform this same analysis on MMSE results due to a ceiling effect in MMSE score distribution suggesting that the MMSE may not be appropriate for highly educated, aging populations.

The composite score created from Feature Match, Odd One Out, and Spatial Planning was highly correlated with MoCA scores (*r*=0.74), indicating that these three tests may be an effective way to track cognitive changes in aging adults, independent of the MoCA.

Due to differences in study execution, we were not able to replicate the results of Brenkel et al. (2017) who found that the Odd One Out and Double Trouble tests best categorized borderline MoCA participants. First, Brenkel et al (2017) used a cut-off score of 27 rather than the score of 26 suggested by MoCA test developers and also used in this study. Second, the participant populations were quite different. In this experiment, participants were highly educated and recruited from retirement homes. Brenkel et al. (2017) recruited from a geriatric psychiatry outpatient clinic and included participants with known mood or major neurocognitive disorders. Finally, our participants completed the tasks on a touch screen tablet computer (iPad) rather than with a mouse and computer screen.

**Conclusions**

This study also suggests that an online test battery is feasible to use with older adults. By 2036, 25% of the Canadian population will be over 65 years of age (Canada, 2016), and the ability to easily assess these individuals will be increasingly important. The CBS battery is ideal because it can be administered without a one-on-one interview, reducing administrator burden, and produces novel test versions each time it is administered, reducing potential practice effects. Moreover, the addition of a single CBS test to the MoCA better identifies individuals with ambiguous scores, and a short (under 10 minutes) battery of just three CBS tests is a viable alternative to the MoCA or MMSE for monitoring cognitive changes in older adults. Future studies will use large samples of older adults with known diagnoses to define thresholds for this novel testing battery in populations with a range of age-related conditions.

**Conflict of Interest Statement**

The online cognitive tests used in this study are marketed by Cambridge Brain Sciences Inc, of which Dr. Owen is the Chief Scientific Officer. Under the terms of the existing licensing agreement, Dr. Owen and his collaborators are free to use the platform at no cost for their scientific studies and such research projects neither contribute to, nor are influenced by, the activities of the company. As such, there is no overlap between the current study and the activities of Cambridge Brain Sciences Inc, nor was there any cost to the authors, funding bodies or participants who were involved in the study.

**References**

Brenkel, M., Shulman, K., Hazan, E., Herrmann, N., & Owen, A. M. (2017). Assessing Capacity in the Elderly: Comparing the MoCA with a Novel Computerized Battery of Executive Function. *Dementia and Geriatric Cognitive Disorders Extra*, *7*(2), 249–256. http://doi.org/10.1159/000478008

Canada, S. (2016). *Census of Population, 1851 to 2016*.

Damian, A. M., Jacobson, S. A., Hentz, J. G., Belden, C. M., Shill, H. A., Sabbagh, M. N., … Adler, C. H. (2011). The montreal cognitive assessment and the mini-mental state examination as screening instruments for cognitive impairment: Item analyses and threshold scores. *Dementia and Geriatric Cognitive Disorders*, *31*(2), 126–131. http://doi.org/10.1159/000323867

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. http://doi.org/10.1016/0022-3956(75)90026-6

Gluhm, S., Goldstein, J., Loc, K., Colt, A., Liew, C. Van, & Corey-Bloom, J. (2013). Cognitive Performance on the Mini-Mental State Examination and the Montreal Cognitive Assessment Across the Healthy Adult Lifespan. *Cognitive Behavioural Neurology*, *26*(1), 1–5. http://doi.org/10.1097/WNN.0b013e31828b7d26.Cognitive

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating Human Intelligence. *Neuron*, *76*(6), 1225–1237. http://doi.org/10.1016/j.neuron.2012.06.022

Malek-Ahmadi, M., Powell, J. J., Belden, C. M., O’Connor, K., Evans, L., Coon, D. W., & Nieri, W. (2015). Age- and education-adjusted normative data for the Montreal Cognitive Assessment (MoCA) in older adults age 70–99. *Aging, Neuropsychology, and Cognition*, *22*(6), 755–761. http://doi.org/10.1080/13825585.2015.1041449

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*, *53*(4), 695–699. http://doi.org/10.1111/j.1532-5415.2005.53221.x

Nasreddine, Z. S., Phillips, N., & Chertkow, H. (2012). Normative data for the montreal cognitive assessment (MOCA) in a population-based sample. *Neurology*, *78*(10), 765–766. http://doi.org/10.1212/01.wnl.0000413072.54070.a3

**Table and Figure Captions**

Table 1. *Task score summary*

Figure 1. MoCA and MMSE scores are plotted against the CBS 3-test composite score with horizontal lines indicating the thresholds used to differentiate the three groups. MoCA scores were categorized using the method described in the Results. MMSE scores were categorized using the severity method described in the published MMSE scoring document. Diagonal lines indicate the correlation between the MMSE (left) or MoCA (right) and CBS 3-test composite scores (significant at p<0.001).

Table 1. *Task score summary*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Task** | **Scores** | | **Population norms**  **age 70-94** | |
|  | **Mean** | **SD** | **Mean** | **SD** |
| MoCA | 24.6 | 4.0 |  |  |
| MMSE | 27.7 | 2.8 |  |  |
| Double Trouble (CBS) | 9.4 | 11.9 | 17.8 | 11.2 |
| Odd One Out (CBS) | 11.4 | 3.5 | 13.4 | 2.4 |
| Spatial Planning (CBS) | 12.4 | 7.1 | 14.4 | 7.5 |
| Grammatical Reasoning (CBS) | 10.4 | 5.9 | 13.8 | 4.6 |
| Digit Span (CBS) | 4.8 | 1.8 | 6.8 | 1.6 |
| Token Search (CBS) | 5.8 | 1.7 | 6.3 | 1.9 |
| Paired Associates (CBS) | 3.5 | 0.9 | 4.3 | 1.0 |
| Spatial Span (CBS) | 4.3 | 1.2 | 4.9 | 0.9 |
| Feature Match (CBS) | 68.3 | 25.6 | 95.8 | 24.8 |
| Rotations (CBS) | 32.4 | 26.1 | 62.4 | 28.6 |
| Polygons (CBS) | 17.7 | 16.8 | 32.4 | 19.9 |
| Monkey Ladder (CBS) | 5.7 | 1.8 | 6.6 | 1.3 |

Summary of task scores for the 50 participants included in this study and relevant population norms from 342 older adults aged 70-94 acquired from the CBS Inc. database of 70 000 participants. In the current study, only 7 participants were younger than 70. For details about the named CBS tests, see Supplementary Materials.

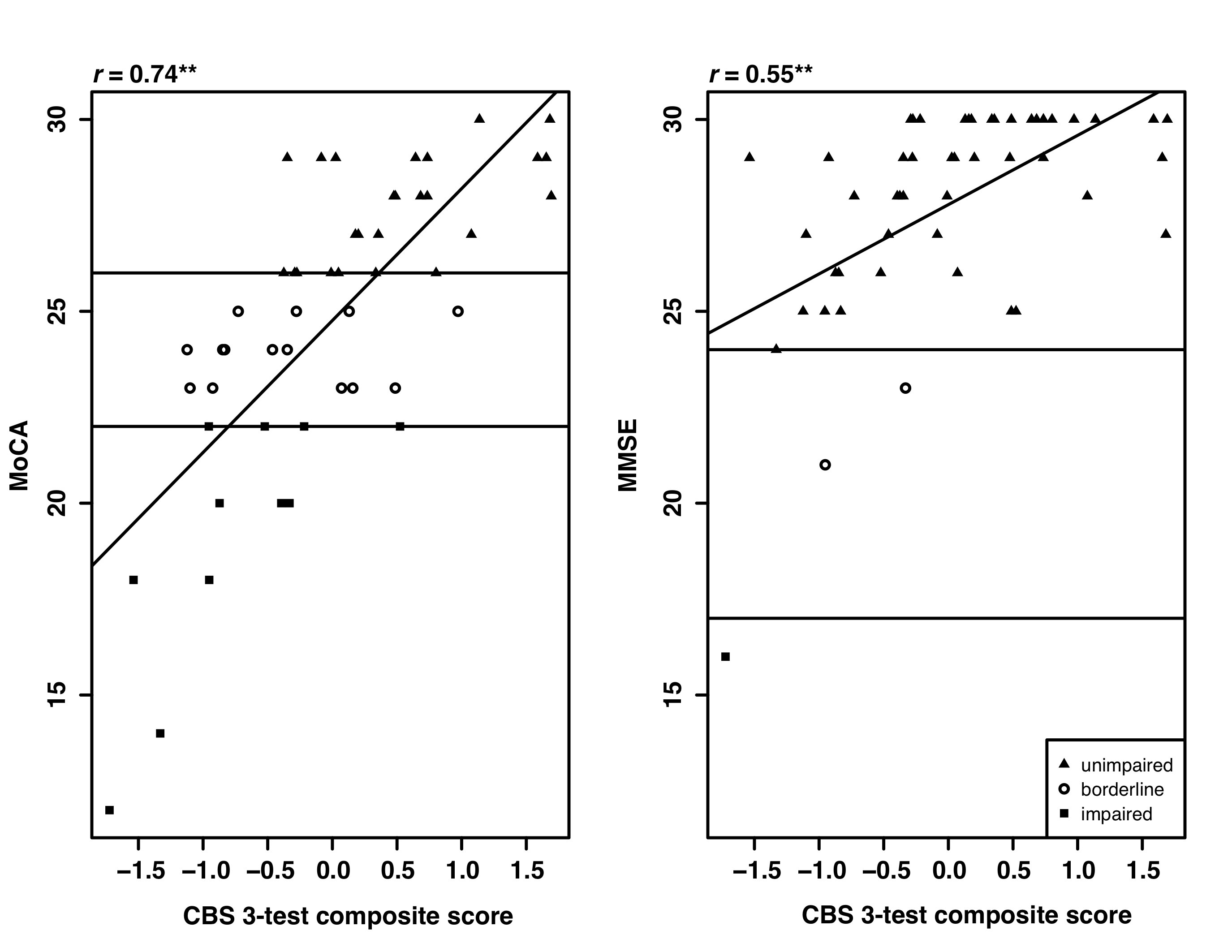


Figure 1. MoCA and MMSE scores are plotted against the CBS 3-test composite score with horizontal lines indicating the thresholds used to differentiate the three groups. MoCA scores were categorized using the method described in the Results. MMSE scores were categorized using the severity method described in the published MMSE scoring document. Diagonal lines indicate the correlation between the MMSE (left) or MoCA (right) and CBS 3-test composite scores (significant at p<0.001).

Supplementary Materials

**Brief Descriptions of the 12 Cambridge Brain Sciences Tasks**

**www.cambridgebrainsciences.com**

|  |  |
| --- | --- |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:DoubleTrouble.pdf | **Double Trouble (Colour-Word Remapping Task)**  A variant on the Stroop test (Stroop, 1935). Three coloured words are displayed on the screen: one at the top and two at the bottom. Participants must indicate which of two coloured words at the bottom of the screen correctly describes the colour that the word at the top of the screen is written in. The colour word mappings may be congruent, incongruent, or doubly incongruent, depending on whether or not the colour of the top word matches the colour that it is written in. Participants have 90 seconds to solve as many problems as possible. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:OddOneOut.pdf | **Odd One Out**  Based on a sub-set of problems from the Cattell Culture Fair Intelligence Test (Cattell, 1949). Nine patterns will appear on the screen. The features that make up the patterns are colour, shape, and number and are related to each other according to a set of rules. Participants must deduce the rules that relate the object features and select the pattern that do not correspond to those rules. Difficulty is increased or decreased depending on whether the participant got the previous trial correct. Participants have 3 minutes to solve as many problems as possible. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:DigitSpan.pdf | **Digit Span Task**  A variant on the verbal working memory component of the WAIS-R intelligent test (Weschler, 1981). A sequence of numbers will appear on the screen one after another. Once the sequence is complete, participants must repeat the sequence by entering them on the keyboard. Difficulty is increased or decreased by one number depending on whether the participant got the previous trial correct. After three errors, the task will end. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:FeatureMatch.pdf | **Feature Match Task**  Based on the classical feature search tasks that have been used to measure attentional processing (Treisman & Gelade, 1980). Two grids are displayed on the screen, each containing an array of abstract shapes. In half of the trials the grids differ by just one shape. Participants must indicate whether or not the grid’s contents are identical. Difficulty is increased or decreased by one shape depending on whether the participant got the previous trial correct. Participants have 90 seconds to solve as many problems as possible. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:Polygons.pdf | **Polygons Task**  Based on the Interlocking Pentagons Task, which is often used in the assessment of age related disorders (Folstein et al., 1975). A pair of overlapping polygons is displayed on one side of the screen. Participants must indicate whether a polygon displayed on the other side of the screen is identical to one of the interlocking polygons. Difficulty is increased by making the differences between the polygons more subtle or decreased by making the differences between the polygons more pronounced. Participants have 90 seconds to solve as many problems as possible. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:PairedAssociates.png | **Paired Associates Task**  A variant on a paradigm that is commonly used to assess memory impairments in aging clinical populations (Gould et al., 2005). Boxes are displayed at random locations on the screen. The boxes are opened one after another to reveal an enclosed object. Subsequently, the objects are displayed in random order in the centre of the screen and participants must determine which box contains the object that is presented. Difficulty is increased or decreased by one box depending on whether the participant got the previous trial correct. After three errors, the task will end. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:TokenSearch.pdf | **Token Search Task**  Based on a test that is used to measure strategy during search behaviours (Collins et al., 1998). Sets of boxes are displayed on the screen in random locations. Participants must find a hidden “token” by clicking on the boxes one at a time to reveal their contents. When the token is found, it is hidden within another box. On any given trial, the token will not appear within the same box twice, thus, participants must search the boxes until the token has been found once within each box. If they search the same empty box twice whilst looking for the token, or search a box in which the token has previously been found, this is an error and the trial ends. Difficulty is increased or decreased by one box depending on whether the participant got the previous trial correct. After three errors, the task will end. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:SpatialTree.pdf | **Spatial Planning Task**  A variant on the Tower of London Task (Shallice, 1982), which is used to measure executive function. Numbered beads are positioned on a tree shaped frame. Participants must repositions the beads so that they are configured in ascending numerical order running from left to right and top to bottom of the tree, in as few moves as possible. Problems become progressively harder with the total number of moves required and the planning complexity increasing in steps. Trials are aborted if the participant makes more than twice the number of moves required to solve the problem. Participants have 3 minutes to solve as many problems as possible. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:Rotations-2.pdf | **Rotations Task**  Often used for measuring the ability to manipulate objects spatially in mind (Silverman et al., 2000). Two grids of coloured squared are displayed to either side of the screen with one of the grids rotated by a multiple of 90 degrees. When rotated, the grids are either identical or differ by the position of just one square. Participants must indicate whether or not the grids are identical. Participants have 90 seconds to solve as many problems as possible. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:SpatialSpan.pdf | **Spatial Span Task**  A variant on the Corsi Block Tapping Task (Corsi, 1972), used for measuring spatial short-term memory capacity. 16 squares are displayed in a 4 x 4 grid. A sub-set of the squares will flash in a random sequence at a rate of 1 flash every 900 ms. Subsequently, participants must repeat the sequence by clicking on the squares in the same order in which they flashed. Difficulty is increased or decreased by one box depending on whether the participant got the previous trial correct. After three errors, the task will end. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:GrammaticalReasoning.pdf | **Grammatical Reasoning Task**  Based on Alan Baddeley’s three minute grammatical reasoning test (Baddeley, 1968). Short sentences describing the relationship of two shapes along with an image of the shapes are displayed on the screen. Participants must indicate whether the sentence correctly describes the pair of objects displayed on the screen. Participants have 90 seconds to solve as many problems as possible. |
| Macintosh HD:Users:avitalsternin:Documents:Western:Academics:PhDProject.git:CBS:Figures:MonkeyLadder.pdf | **Monkey Ladder (Visuospatial Working Memory Task)**  A variant on a task from the non-human primate literature (Inoue & Matsuzawa, 2007). Sets of numbered squares are displayed on the screen at random locations. After a variable interval of time, the numbers disappear leaving just the blank squares and participants must respond by clicking the squares in ascending numerical sequence. Difficulty is increased or decreased by one numbered box depending on whether the participant got the previous trial correct. After three errors, the task will end. |

Supplementary Figure 1



FigureS1*.* As part of a large online study 327 participants (age 71-80 years) completed all 12 CBS tasks. The standardized scores are presented in this figure and show relative task difficulty. The three tests in bold (Feature Match, Odd One Out, Spatial Planning) are those identified by the analyses in the current paper. As described in the main text, task difficulty can not account for the effects observed.